JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1998, **31,** 387–398 NUMBER 3 (FALL 1998)

DEGREE OF CONSTRUCTED-RESPONSE INTERACTION IN COMPUTER-BASED PROGRAMMED INSTRUCTION

KALE M. KRITCH

ASSOCIATED MARINE INSTITUTES

AND

DARREL E. BOSTOW

UNIVERSITY OF SOUTH FLORIDA

This investigation evaluated the importance of frequent responding in computer-based programmed instruction. Instructional computer programs that taught the use of an authoring language were administered to 155 undergraduate college students. One group experienced frequent (dense) situations requiring them to supply key components of the subject taught. A second group experienced half as many response requirements, and a third "passive" group simply tapped any key to progress. To control for time in contact with presentations, individuals in a fourth group were yoked to the members of the high-density requirement group. Statistically significant differences on both posttest and application performances indicated that students who experienced the high density of overt response contingencies scored the best and the passive group score the worst. The yoked control revealed that time on task alone could not account for the superior performance of students in the high-density group. Results suggest that inclusion of a high rate of constructed-response contingencies within instructional computer programs increases performance.

DESCRIPTORS: programmed instruction, computer-based instruction, interaction, constructed response, contingency

---

Programmed instruction conforms closely to what has been learned in the operant conditioning laboratory, especially to the features of the contingency of reinforcement. Skinner (1963) defined programming as "the construction of carefully arranged sequences of contingencies leading to the terminal performances which are the object of education" (p. 183).

Holland (1967) pointed out that a key concept in this definition is the contingency. Instructional materials are designed in such a way that correct answers are achievable only after some precursory behavior. Only the precursory behavior that had to occur to obtain a correct answer is assured, and only that unavoidable behavior constitutes the contingency for a correct answer. Material is, then, programmed only to the extent that the correct answer is contingent upon appropriate preceding behavior.

However, as Vargas and Vargas (1992) later pointed out, Holland's definition of programmed instruction was not widely understood early in the programmed instruction movement. Many, if not most, examples of programmed instruction were not programmed according to the above definition. Summarizing the situation 30 years ago, Holland (1967) wrote,

We thank Ann Barron, Louis E. Bowers, Andria M. Troutman, and James A. White for guidance and support throughout this study. This study is based on Kale M. Kritch's doctoral dissertation within an interdisciplinary program of behavior analysis and instructional technology.

Correspondence concerning this article should be addressed to Kale M. Kritch, who is now the Director of Research at the Associated Marine Institutes, Inc., 5915 Benjamin Center Drive, Tampa, Florida 33634 (E-mail: kkritch@ami-fl.org).

> Programmed instruction, at its best, is an applied science in that it rests on fundamental principles of the learning laboratory. However, the principles as they have been related to matters of program design have been all too vague and general. They have lost much of the precision of operational statements so necessary to experimental research. The loss is unfortunate because exact meanings and empirically identifiable processes distinguish technology from craft. Further development of a technology of programmed instruction awaits the return to exact specifications for the important variables. It is the task of applied research to sharpen the description of fundamental variables needed in application and to forge useful measuring instruments for these variables. (p. 87)

Unfortunately, the stricter definition of programmed instruction appears to be rarely understood today, especially in the field of instructional technology. Although critical components of programmed instruction were isolated long ago (Doran & Holland, 1971; Holland 1964, 1965, 1967; Holland & Kemp, 1965; Skinner, 1963), well-constructed programs have yet to be widely applied (Vargas & Vargas, 1992). Researchers who evaluate instructional programs often appear to be unaware of or to misunderstand the learning principles and their implications for effective instructional technology.

To confirm this suspicion, the present authors closely inspected five widely recognized meta-analyses in the computer-based instruction (CBI) field (Bangert-Downs, Kulik, & Kulik, 1985; Becker, 1988; C. Kulik & Kulik, 1991; J. A. Kulik, Kulik, & Bangert-Downs, 1985; J. A. Kulik, Kulik, & Cohen, 1980) for mention of essential techniques and key concepts of programmed instruction. This inspection revealed that the meta-analyses incorporated different types of instructional programs, including tutorial, management, general enrichment, drill and practice, programming, and simulation programs. Less than 2% of these manuscripts specified that the tutorials contained text-based stimuli. Less than 20% listed constructed response as the behavior required of the student. Less than 13% contained any evidence of differential reinforcement. Less than 2% of the manuscripts mentioned the concept of prompting. No manuscript mentioned the processes of fading, shaping, or priming. These findings suggest the inappropriateness of comparing experimental results of computer-based programmed instruction with traditional CBI and computer-assisted instruction (CAI).

Research in the area of CBI that most closely relates to the notion of instructional contingencies includes investigations of interactivity or, more specifically, embedded questions. Research results in the area of videodisc interactivity suggest a positive relationship between level of interactivity and student achievement (Fletcher, 1990; Schaffer & Hannafin, 1986).

Similarly, Hannafin (1985) concluded that (a) the type and nature of interactivity affected the magnitude and the nature of learning, (b) learning was most effective when embedded items did not follow excessively long segments, and (c) comprehension was increased by response feedback. Lofald and Pajares (1993) reported that embedded questions (in expository text on the computer) increased achievement more than electronic page turning through the same material. (However, student responding included only a single multiple-choice item after every fifth electronic page of material.) Dalton and Hannafin (1987) included practice items in video-based materials and found that actively responding to embedded questions resulted in greater achievement scores than viewing videotapes without practice questions.

Although results from studies such as these suggest a positive relationship between responding to embedded questions and achievement, a study by Phillips, Hannafin, and Tripp (1988) showed that increased levels of interaction from nonexistent to limited to elaborative embedded questions had no relationship to amount of learning.

More recently, a number of behavior analysts have turned their attention to the verification of the importance of the overt constructed student response and its role in achievement of skills. Using a pretest-posttest design, Thomas and Bostow (1991) compared unprogrammed paper prose, segmented prose, program prose with blanks, and constructed-response computer-programmed instruction. Results revealed that overt responding (paper and computer) produced significantly better posttest performance, and constructed-response interaction required significantly more time to complete. Similar results with different content (Kumar, Bostow, Schapira, & Kritch, 1993; Tudor & Bostow, 1991) and materials (videodisc instruction; Kritch, Bostow, & Dedrick, 1995) have been reported in the literature.

Overt student involvement during tutorials appears to be important. But does the extent of what is learned vary in a consistent way with the density or frequency of required constructed responses? The present study examined the relations between three levels of the density of constructed-response contingencies within computer-programmed instructional tutorials and two outcomes: (a) achievement gains measured by a verbal posttest and (b) the extent to which students could subsequently accomplish a relevant applied skill. Because it might be argued that time on task alone could be responsible for differences in effectiveness, the study included a yoked control.

METHOD

Participants

Undergraduate education majors (N = 155) from an educational foundations course at the University of South Florida served as participants. Of these, 70% were female. The average age was 25 years (SD = 0.46) and ranged from 19 to 60 years. Participants had the option of participating in the study or writing a term paper; however, none chose to write the term paper. Students were randomly assigned to experimental conditions by a computer program.

Materials

The computer laboratory contained 12 IBM 8088 Model XT computers that were used to deliver the instructional programs. Materials unrelated to the study were removed. The computers were placed on tables against the walls of the laboratory with large cardboard dividers forming work stations.

The instructional programs were constructed using the PC-CAI Authoring System, version 2.10. Instructional design principles and techniques prescribed for computer-based programmed instruction in the program called Creating Computer Programmed Instruction (Kritch & Bostow, 1994) were used to create the instructional software programs.

In one experimental condition, computer operation required connecting pairs of computers via null modem cables. Two programs written in the BASIC programming language allowed connected computers to work synergistically.

Treatment Conditions

A 176-frame instructional program about PC-CAI programming commands was constructed prior to conducting the present study. This program was field tested and revised using data from 28 students drawn from a graduate course in learning and instructional design. The program was cast into the following formats for experimental evaluation of their different effects upon student performance.

High-density (HD) program. The HD constructed-response condition contained 176 instructional contingencies, each providing a screen of instructional material (i.e., a "frame") and a blank to be filled in by typing an overt constructed-response word at the keyboard. Because of programming limitations, the HD program also included six traditional multiple-choice items dispersed throughout the program in which at least two alternatives were presented.

When a student typed the correct response, the computer displayed "Correct" in the center of the screen and then presented the next frame. If the answer given by the student was incorrect, the computer displayed "Incorrect" and the correct answer, and instructed the student to press any key to continue to the next frame.

Low-density (LD) program. A second version of the same content material was constructed. This LD constructed-response program contained 88 instructional contingencies, or exactly 50% of the HD version. The LD condition was created by filling in blanks with the correct response in every other frame of the HD condition. Similar to the HD condition, students in this condition were also required to construct responses to frames, and answers were handled the same way. Thus, one frame required a constructed response, the next simply a key press.

Zero-density (ZD) condition. A third version, the ZD constructed-response condition, contained no overt constructed-response contingencies; all blanks were filled in. Students passively read each instructional frame in the same linear order but simply tapped any key to continue.

Yoked control (for time) (CT) condition. This condition was identical in content to the ZD condition. However, students in the CT condition were randomly paired with students in the HD condition and were required to experience each instructional frame for the same length of time as the paired participants. Here, as the yoked participants in the HD program entered a response and progressed to the next frame, the frame that the paired participant in the CT condition experienced was automatically advanced to the next frame. Therefore, students in the CT condition experienced each identical instructional frame for the same length of time as his or her paired cohort in the HD condition. The only difference between these conditions was the presence and absence of blanks with the necessity of actively responding.

Dependent Variables

Various demographic details were collected in the form of a questionnaire at the beginning of the course. In addition, each version of the instructional program automatically recorded participants' names, individual responses, percentage correct scores, and time taken to complete the program.

A 34-item posttest evaluated the degree of verbal knowledge acquired by students in the various treatment conditions. Each posttest item described an effect to be achieved and called for the programming code necessary to achieve the described effect.

Another dependent variable was the quality of a student product. This application of knowledge required the student to write on a piece of graph paper the series of PC-CAI program commands necessary to present a single interactive frame.

Procedure

One week before the intervention, each student scheduled a 2-hr appointment for a "special event" at the computer laboratory. Students were not informed about the nature of the study. Six students missed their appointments due to illness. Computer malfunctions caused the loss of experimental data from 3 students. Four students were excused during the experiment because of illness. This attrition resulted in slightly unequal groups (HD = 42, LD = 38, ZD = 39, CT = 36).

After arriving at the laboratory, students were ushered by the laboratory manager to a randomly assigned computer station, and each was instructed to read identical directions on paper. Questionnaires had revealed that no student reported familiarity with the PC-CAI language, and for this reason a pretest was not included. The first author constantly monitored the computer laboratory throughout the experiment.

After completing the instructional program, the laboratory manager presented the student with a sheet of graph paper, pencil, and directions that asked each student to construct as many of the necessary PC-CAI commands in their proper sequence for creating an instructional frame as he or she could. Graph paper was used to permit accurate scoring of the character placement of commands.

After completing the application test, materials were removed and students experienced the computer posttest. The computer posttest recorded each response, the time taken to complete each item, and the percentage correct score for each student. Students were not informed of their posttest scores (i.e., either the application or computer test) to minimize postexperiment discussion.

The Kuder-Richardson 20 (Borg & Gall, 1989) test for internal reliability was calculated post hoc and yielded a reliability coefficient of .91 for the computer posttest. To determine interrater reliability for scoring the application posttest, the service of a graduate assistant who was unfamiliar with the PC-CAI authoring program and code was enlisted. After working through the HD program twice, the graduate assistant scored a random sample of 30 application products using the product grade sheet and key. Her scores were compared with those of the author, who scored application products using the identical product grade sheet, and 100% agreement occurred.

The Kuder-Richardson 20 test for internal reliability was also calculated post hoc for items contained on the application grade sheet, and a reliability coefficient of .96 was obtained.

Upon completion of the computer posttest, students completed a postexperiment questionnaire. This questionnaire briefly explained the nature of the experiment and assessed student attitudes regarding the experiment and computer instruction.

Because appointments were scheduled at the same location throughout the week, it was anticipated that discussion between students might occur. Therefore, each student was given a debriefing contract immediately after completing the final questionnaire that described the importance of conducting educational research and asked each student to sign a pledge not to speak to anyone about the experiment until the results were provided by the instructors. All students willingly signed the debriefing contract.

Experimental Design and Data Analysis

A one-way analysis of variance (ANOVA) was employed to evaluate differences in posttest scores and student-written application products (Borg & Gall, 1989). An additional 4 (Instructional Groups) $\times$ 2 (Ability) ANOVA consisting of two between-groups factors was employed to evaluate whether posttest and application performances varied with respect to ability level, as measured by current grade-point averages, across instructional conditions. Pearson product-moment correlations were calculated to describe the strength of relationships between variables.

In all statistical comparisons, an alpha level of .05 was applied.

Analysis of variance revealed that random assignment produced groups that were not significantly different with respect to the distribution of grade-point average (GPA), age, and gender. Data records were assembled into summary charts used for the SAS statistical program (Cody & Smith, 1991). Figure 1 summarizes the experimental conditions and response contingencies, and provides example instructional frames, test content, and questionnaire items presented to the four groups.

RESULTS

Results of the ANOVA on computer posttest scores revealed significant differences between groups, $F(3, 151) = 24.44$, $p < .0001$. Table 1 presents ANOVA results, and Table 2 presents posttest means of the computer posttest, application test, and time to complete the instructional programs for the four instructional conditions. Results revealed a nearly 30-point difference between the HD and ZD groups.

The top panel of Figure 2 shows that there was a positive relationship between the density of constructed-response contingencies and the students' performances on the computer posttest. The HD group had the highest mean score, followed next by the LD group, then the CT group, and finally the ZD group. Pairwise post hoc comparisons using Tukey's test indicated significant differences between the HD and CT groups, the HD and ZD groups, the LD and CT groups, and the LD and ZD groups. The more frequently contingencies required overt constructed responses, the higher the posttest percentage correct scores. Requiring control subjects to spend time equal to a yoked high-density cohort failed to produce similar posttest performance.

Results of the ANOVA calculated from application posttest scores revealed significant differences between groups, $F(3, 140) = 3.96$, $p < .0095$. As presented in the middle panel of Figure 2, pairwise post hoc comparisons using Tukey's test indicated significant differences between the HD and ZD groups. ANOVA results with regard to the time taken to complete each version of the instructional program revealed significant differences between groups, $F(3, 151) = 50.72$, $p < .0001$.

As presented in the bottom panel of Figure 2, pairwise post hoc comparisons using Tukey's test indicated significant differences in time taken between all groups except between the HD and CT groups. Because the latter groups were yoked, one could not expect them to differ. Although each CT participant experienced the instructional program for the same length of time as each paired HD participant, time differences between these groups resulted from the subject attrition described previously.

ANOVA results using the time taken to complete the computer posttest also revealed significant differences between groups, $F(3, 151) = 6.81$, $p < .0002$. There was a negative relationship between the number of minutes taken to complete the posttest and the density of constructed-response contingencies. Pairwise post hoc comparisons using Tukey's test on the time taken for the computer posttest indicated significant differences between the HD and ZD groups and the HD and CT groups. ANOVA results using the time taken to complete the application test revealed significant differences between groups, $F(3, 151) = 7.47$, $p < .0001$.

Pairwise post hoc comparisons using Tukey's test indicated significant differences in time taken on the application test between the HD and LD groups, the LD and CT groups, and the ZD and CT groups.

Pearson correlation coefficients were calculated on variables at the $p < .05$ level, and statistically significant correlations were

High Density (HD) Overt responses to 176 frames	**Low Density (LD)** Overt responses to every other frame	**Zero Density (ZD)** Passive reading, key tapping to advance	**Control for Time (CT)** Passive reading, advance when HD advanced
1. A player piano is told what notes to play from a long scroll of paper with tiny holes punched through it. The paper scroll is like a script of commands that tells the piano what n____s to play.	1. A player piano is told what notes to play from a long scroll of paper with tiny holes punched through it. The paper scroll is like a script of commands that tells the piano what n____s to play.	1. A player piano is told what notes to play from a long scroll of paper with tiny holes punched through it. The paper scroll is like a script of commands that tells the piano what notes to play.	1. A player piano is told what notes to play from a long scroll of paper with tiny holes punched through it. The paper scroll is like a script of commands that tells the piano what notes to play.
2. With a player piano, the music is programmed. The scroll of paper is a script of commands that tells the player _____ what notes to play.	2. With a player piano, the music is programmed. The scroll of paper is a script of commands that tells the player piano what notes to play.	2. With a player piano, the music is programmed. The scroll of paper is a script of commands that tells the player piano what notes to play.	2. With a player piano, the music is programmed. The scroll of paper is a script of commands that tells the player piano what notes to play.
3. Like a player piano, a computer can be programmed. A computer program is like a sc____pt of commands that tells the c____r what to do.	3. Like a player piano, a computer can be programmed. A computer program is like a sc____pt of commands that tells the c____r what to do.	3. Like a player piano, a computer can be programmed. A computer program is like a script of commands that tells the computer what to do.	3. Like a player piano, a computer can be programmed. A computer program is like a script of commands that tells the computer what to do.

Yoked

Sample Test Items
1. The "statements" that cause a computer program to take actions are called ______.
2. The command that erases any previous material from the screen is the ______ command.
3. The command that tells the program to start a new frame is the ______ ______ command.

Questionnaire Items (1-very much dislike; 2=dislike; 3=neutral; 4=like; 5=very much like)
How would you describe your "attitude" about the instructional program that you experienced today?
How would you describe you "attitude" about computer assisted instructional programs in general?
How would you describe your "attitude" about computer assisted instructional programs that specifically teach program commands like those taught in the instructional program you just experienced?

Figure 1. Summary of conditions, sample frames, test, and questionnaire items.

found. For the two groups that produced tutorial percentage correct scores (i.e., the HD and LD groups), tutorial percentage correct scores positively correlated with posttest percentage correct scores (r = .39 and r = .81, respectively), and tutorial percentage correct scores positively correlated with application percentage correct scores (r = .35 and r =

Table 1
ANOVA Results on the Computer Posttest, Application Test, and Time to Complete the Instructional Programs for the Four Groups

Source	*df*	Sum of squares	Mean square	*F* value	$p > F$
Computer posttest					
Model	3	20,276.77	6,758.92	24.44	0.0001
Error	151	41,753.80	276.51		
Total	154	62,030.57			
Application test					
Model	3	0.4278	0.143	3.96	0.0095
Error	140	5.0392	0.036		
Total	143	5.4669			
Time to complete programs					
Model	3	34,024.20	11,341.40	50.72	0.0001
Error	151	33,763.38	223.60		
Total	154	67,787.57			

.41, respectively). Also, time to complete the tutorial positively correlated with time to complete the posttest within both groups ($r = .34$ and $r = .53$, respectively). Within the HD, LD, ZD, and CT conditions, posttest percentage correct scores positively correlated with application percentage correct scores ($r = .62$, $r = .60$, $r = .69$, and $r = .61$, respectively).

To evaluate whether posttest and application performances varied with respect to ability level (as measured by current GPAs) across instructional conditions, a 4 (Density of Constructed-Response Contingencies) × 2 (Ability) ANOVA consisting of two between-groups factors was used. An alpha level of .05 was applied. Tukey's post hoc multiple comparison test was used to evaluate differences between means. The ability factor was determined by ranking students from low to high based upon their self-reported current GPAs. The highest GPA was 4.0, the lowest was 2.12, and the average was 3.08 ($SD = 0.48$). The median score of 3.0 was calculated, and students whose GPA resided below the median score were placed into the low-ability group; students whose GPAs resided above the median score were placed in the high-ability group. Results of the 4 × 2 ANOVA revealed no significant interaction effect for posttest scores, $F(3, 143) = 1.58$, $p < .197$, or for application scores, $F(3, 135) = 1.75$, $p < .161$. Thus, the effects brought about by the increased density of constructed responding were not shown to be different for students of low ability compared to those of high ability.

ANOVA results on Likert-type responses to final questionnaire items revealed significant differences between groups on only the first question. Analysis of responses to Ques-

Table 2
Means on the Computer Posttest, Application Test, and Time (in Minutes) to Complete the Instructional Programs for the Four Groups

Group	*N*	*M*	*SD*
Computer posttest			
HD	42	85.23	8.57
LD	38	75.75	16.51
ZD	39	56.05	19.83
CT	36	63.09	19.81
Application test			
HD	37	0.767	0.163
LD	37	0.676	0.186
ZD	36	0.616	0.218
CT	34	0.696	0.190
Time to complete programs			
HD	42	80.43	16.50
LD	38	59.03	12.71
ZD	39	48.05	13.19
CT	36	83.78	16.92

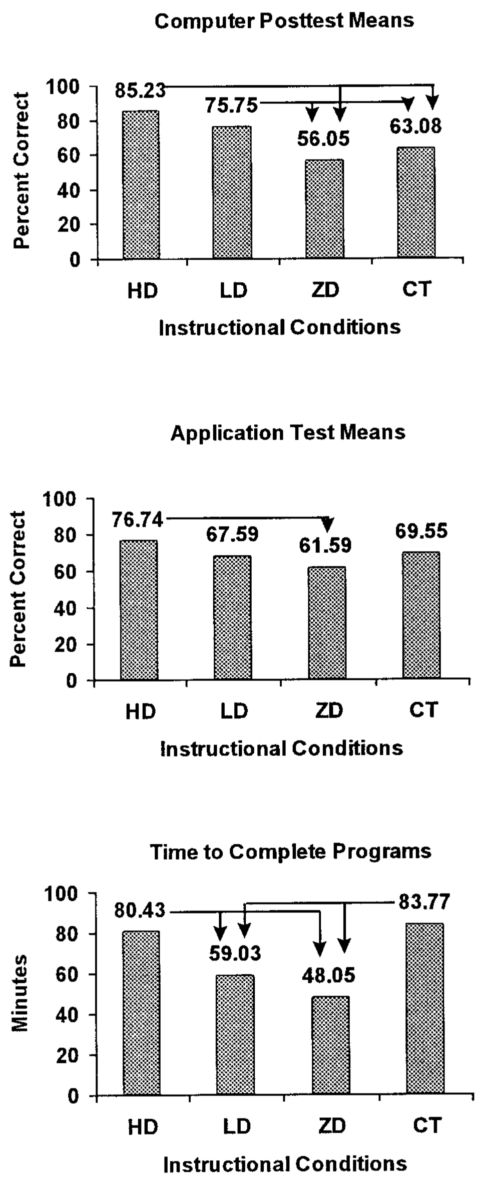

Figure 2. Means for computer posttest, application test, and time taken to complete instructional programs. Arrows indicate statistically significant differences. HD = high-density condition; LD = low-density condition; ZD = zero-density condition; CT = yoked control (for time) condition.

tion 1 ("How would you describe your 'attitude' about the instructional program that you experienced today?") revealed significant differences between groups, $F(3, 155) = 7.78$, $p < .0001$. The HD group produced the most positive mean ($M = 3.43$, $SD = 1.11$), then the LD group ($M = 3.21$, $SD = 0.99$), the ZD group ($M = 2.56$, $SD = 0.95$), and the CT group ($M = 2.54$, $SD = 1.10$). Pairwise post hoc comparisons using Tukey's test indicated significant differences between the HD and ZD groups, between the HD and CT groups, between the LD and ZD groups, and between the LD and CT groups. In other words, a positive relationship between participants' attitudes about the instructional program they experienced and the density of constructed-response contingencies was revealed. The more they had to respond, the better they liked doing the program.

DISCUSSION

This study extended the existing literature by investigating the functional relations between varying densities of constructed-response contingencies with respect to two outcomes: (a) achievement gains measured by a computer-delivered posttest, and (b) the extent to which students could subsequently accomplish a relevant applied skill (writing program code). The study included a condition to control for time on task, and also searched for possible interaction effects between student ability levels and treatment conditions.

In the present study, the increased density of constructed-response contingencies produced increased strengthening of intraverbal repertoires, as measured by a computer fill-in-the-blank posttest. These results provided further support for the argument that frequent overt constructed responding within instructional contingencies is a critical design feature for effective computer-based instruction. The statistically significant differences between posttest group means appear large enough to have important practical implications. Requiring more frequent responses takes students more time, but it appears to be the nature of the response requirement (i.e., the response contingencies) and not the time spent that is the source of better subsequent performance. Results in this study highlight the importance of instructional techniques that require learners to repeatedly and overtly engage in components of the desired terminal behavior.

Results from the present study showed that the inclusion of a high density of constructed responding increased not only performance on a quiz resembling the instructional frames but also the extent to which students could subsequently accomplish a relevant applied skill (writing program code). This finding provides further support of previous results obtained by Tudor and Bostow (1991).

Results of the present study support the position that increased interactivity produces increased learning (Fletcher, 1990; Schaffer & Hannafin, 1986). However, these results do not support the conclusion that less time is required when learning from computer-based instruction (J. A. Kulik et al., 1980). Indeed, the present research showed that significantly more instructional time was consumed when learning was greater. The present results were consistent with previous research in the area of computer-programmed instruction (Thomas & Bostow, 1991; Tudor & Bostow, 1991).

In the present study, the correlational analysis revealed a positive relationship between success during instruction and resulting performance. In other words, the lower the error rate during instruction, the better the performance after instruction. Such results support the widely held assumption in the field of programmed instruction that good programs should have low error rates (Holland, Solomon, Doran, & Frezza, 1976).

A past criticism of traditional programmed instruction was that it may be appropriate only for teaching rote memory of basic information to low-ability students, a criticism noted by Vargas and Vargas (1992). However, results of the 4×2 ANOVA revealed no significant interactions. In other words, the effects brought about by the increased density of constructed responding were not different for low- or high-ability students.

Instructional material used in this study almost exclusively consisted of text-based stimuli. Therefore, results from this research may generalize primarily to those situations in which instructional stimuli consist mainly of verbal or textual information. The generality of the effects achieved here needs to be further evaluated with respect to other forms of instructional stimuli.

The instructional programs in this study established mainly intraverbal repertoires consisting of programming language and code. Results from this study may generalize most specifically to those situations in which intraverbal repertoires of programming or scripting languages are the terminal products of instruction. Clearly, research that extends the analysis to other skills should be undertaken.

The present research design incorporated treatments administered only once and to different groups. Summarizing comparisons therefore unavoidably mixed between-subjects differences into the statistical analysis. A single-subject multitreatment design would have required several different instructional programs or sections of programs, each student being subjected to differing sequences of the treatments. Such an experimental program was beyond the scope and resources of the experimenters and would have brought with it a host of complex logistical considerations. Future systematic replications of the present research could, however, adopt single-subject strategies in an effort to determine the degree to which students may be variously sensitive to the density of response requirements. The field remains a fertile area for creative research designs.

The experimental question ("Will the density or number of instructional contingencies directly relate to how much or what is learned?") seems to have been answered. The present study showed that the greater the number of responses required, the higher the achievement gains. Future research may provide a closer examination of the extent to which increased constructed-response contingencies relate to outcome measures by incorporating a finer graded continuum of various densities of contingencies.

There are many alluring aspects of currently available computer hardware and software. But the field of instructional research may benefit more from focusing on the response requirements during instruction rather than on attractive software or hardware features. Results of the present study highlight the importance of frequent response contingencies within an instructional program—contingencies that incorporate elements of targeted terminal performance.

REFERENCES

Bangert-Downs, R. L., Kulik, J. A., & Kulik, C. C. (1985). Effectiveness of computer-based education in secondary schools. *Journal of Computer-Based Instruction, 12,* 59–68.

Becker, H. J. (1988). The effects of computer use on children's learning: Limitations of past research and a working model for new research. *Peabody Journal of Education, 64,* 81–110.

Borg, W., & Gall, M. (1989). *Educational research: An introduction.* New York: Longman.

Cody, R., & Smith, J. (1991). *Applied statistics and the SAS programming language* (3rd ed.). New York: Elsevier.

Dalton, D. W., & Hannafin, M. J. (1987). The effects of knowledge versus content-based strategies on information and application learning from interactive video. *Journal of Computer-Based Instruction, 14,* 138–141.

Doran, J., & Holland, J. G. (1971). Eye movements as a function of the response contingencies measured by blackout technique. *Journal of Applied Behavior Analysis, 4,* 11–17.

Fletcher, J. D. (1990). *Effectiveness and cost of interactive videodisc instruction in defense training and education* (IDA Paper No. P-2372). Alexandria, VA: Institute for Defense Analyses.

Hannafin, M. J. (1985). Empirical issues in the study of computer-assisted interactive video. *Educational Communication and Technology, 33,* 235–247.

Holland, J. G. (1964). Response contingencies in teaching-machine programs. *Journal of Programmed Instruction, 3,* 1–8.

Holland, J. G. (1965). Research on programming variables. In R. Glaser (Ed.), *Teaching machines and programmed learning: Vol. 2. Data and direction* (pp. 264–269). Washington, DC: National Education Association.

Holland, J. G. (1967). A quantitative measure for programmed instruction. *American Educational Research Journal, 4,* 87–101.

Holland, J. G., & Kemp, F. D. (1965). A measure of programming in teaching machine material. *Journal of Educational Psychology, 56,* 264–269.

Holland, J. G., Solomon, C., Doran, J., & Frezza, D. A. (1976). *The analysis of behavior in planning instruction.* Reading, MA: Addison-Wesley.

Kritch, K. M., & Bostow, D. E. (1994). Creating computer programmed instruction [Computer program]. Tampa, FL: Customs Systems International, Inc.

Kritch, K. M., Bostow, D. E., &, Dedrick, R. F. (1995). Level of interactivity of videodisc instruction on college student recall of AIDS information. *Journal of Applied Behavior Analysis, 28,* 85–86.

Kulik, C., & Kulik, J. A. (1991). Effectiveness of computer-based instruction: An updated analysis. *Computers in Human Behavior, 7,* 75–94.

Kulik, J. A., Kulik, C. C., & Bangert-Downs, R. L. (1985). Effectiveness of computer-based education in elementary schools. *Computer in Human Behavior, 1,* 59–74.

Kulik, J. A., Kulik, C. C., & Cohen, P. A. (1980). Effectiveness of computer-based college teaching: A meta-analysis of findings. *Review of Educational Research, 50,* 525–544.

Kumar, N. B., Bostow, D. E., Schapira, D. V., & Kritch, K. M. (1993). Efficacy of interactive, automated programmed instruction in nutrition education for cancer prevention. *Journal of Cancer Education, 8,* 203–211.

Lofald, D. F., & Pajares, J. F. (1993, April). *The effect of embedded questions on readers' calibration of test readiness.* Paper presented at the annual meeting of the American Educational Research Association, Atlanta.

Phillips, T. L., Hannafin, J. J., & Tripp, S. D. (1988). The effects of practice and orienting activities on learning from interactive video. *Educational Communication and Technology, 36,* 93–102.

Schaffer, L., & Hannafin, M. J. (1986). The effects of progressively enriched interaction on learning

from interactive video. *Educational Communication and Technology Journal, 34,* 89–96.
Skinner, B. F. (1963). Reflections on a decade of teaching machines. *Teachers College Record, 65,* 168–177.
Thomas, D. L., & Bostow, D. E. (1991). Evaluation of pre-therapy computer-interactive instruction. *Journal of Computer-Based Instruction, 18,* 66–70.
Tudor, R. M., & Bostow, D. E. (1991). Computer-programmed instruction: The relation of required interaction to practical application. *Journal of Applied Behavior Analysis, 24,* 361–368.
Vargas, E. A., & Vargas, J. S. (1992). Programmed instruction and teaching machines. In R. P. West & L. A. Hamerlynck (Eds.), *Designs for excellence in education: The legacy of B. F. Skinner* (pp. 33–39). Longmont, CO: Sopris West.

Received October 31, 1996
Initial editorial decision January 17, 1998
Final acceptance February 10, 1998
Action Editor, Julie S. Vargas

STUDY QUESTIONS

1. Although the authors did not define the term *constructed response,* what is the difference between constructed and nonconstructed responses, based on information contained in the article?

2. Describe the purpose of the present study.

3. Briefly describe the high-density (HD), low-density (LD), and zero-density (ZD) conditions.

4. What was the yoked control and what was its importance?

5. What did the application task entail, and why was it an important outcome measure?

6. Summarize the performances of participants assigned to the various conditions.

7. What relationship was observed between error rate during instruction and performance during the application test, and what are the potential implications of this finding?

8. The authors acknowledged the need for systematic replication of their findings using single-subject methodology. How might one examine the effects of frequency of constructed responses using such methodology?

Questions prepared by Eileen Roscoe and Rachel Thompson, The University of Florida